

# Finite Blocklength Analysis of Energy Harvesting Channels

K Gautam Shenoy, *Student Member, IEEE*, and Vinod Sharma, *Senior Member, IEEE*

## Abstract

We study DMC and AWGN channels when the transmitter harvests energy from the environment. These can model wireless sensor networks as well as Internet of Things. We study such channels with infinite energy buffer in the finite blocklength regime and provide the corresponding achievability and converse results.

## Index Terms

Achievable rates, Converse, Channel Capacity, Finite Blocklength, EH-AWGN, EH-DMC.

## I. INTRODUCTION

In the information theoretic analysis of channels, channel capacity is the maximum rate at which a source can transmit messages to the receiver subject to an arbitrarily small probability of error. However, channel capacity can be achieved arbitrarily closely by using very large blocklength codes. In practice, we are restricted by blocklength and as a result, we would like to study the backoff from capacity as well as the variation in maximal code size as a function of blocklength.

Like channel capacity, a finite blocklength characterization consists of two parts, namely the achievability and the converse bound on the maximal code size (number of messages)  $M$ . Given the probability of error, the achievability part usually deals with the existence of a code using, for instance, random coding arguments or manipulating general achievability bounds and showing that the bound can be achieved. The converse, on the other hand is an upper bound on the maximal code size which is to be satisfied by every feasible code. This paper focuses on developing both for the energy harvesting channels.

Energy harvesting channels and networks have gained considerable interest recently due to advances in wireless sensor networks and green communications (see [2], [3] and [4]). Transmitting symbols requires energy at the encoder end. Thus the study of the channel is done in tandem with the energy harvesting system. The energy harvesting section is modeled as a buffer or a rechargeable battery which stores incoming energy from some ambient source (e.g. solar energy from the sun). The energy buffer may be of finite or infinite length and the energy arrival process may be discrete or continuous. A problem of interest is to compare the performance of a channel with and without the energy harvesting system (e.g. whether we can quantify the impact on the channel capacity, finite blocklength capacity, etc.).

Finite blocklength analysis for discrete memoryless channels (DMC) was first carried out by Strassen [5]. Hayashi [6] provided non-asymptotic second order results for AWGN channels in addition to other channel types. The results for DMCs and AWGN channels were further refined by Polyanskiy et al. [7], which provided the third order terms in [7] and developed a meta-converse, a converse result that recovered and improved upon known converses. Tighter results for various DMC's were studied by Tomamichel et al. in [8]. Non-asymptotic analysis of channels with state was carried out in [9]. Assuming infinite buffer, the channel capacity for EH-AWGN channels was obtained in [10] and [11]. The study of finite blocklength achievability for energy harvesting noiseless binary channels was carried out in [12]. Non-asymptotic achievability for EH-AWGN channels and EH-DMC's was developed in [13] where the second order term was  $O(\sqrt{n \log n})$ .

In this paper, we provide an achievable scheme for EH-AWGN channels with infinite buffer. The scheme improves over previously known bounds and also provides the optimal second order term (albeit not necessarily the optimal second order coefficient) of  $\sqrt{n}$ . It is shown that a save and transmit scheme where the saving phase is  $O(\sqrt{n})$  long is sufficient to allow for reliable communication in an energy harvesting set up. When compared with the non-energy harvesting case (but with average power constraint), we observe that the second order term is not increased due to energy harvesting. Note that the coefficients of the second order term would not necessarily be same.

Next we provide a finite blocklength converse for energy harvesting channels. As of now, there hasn't been any published work in coming up with a finite blocklength converse for these channels. In this paper, we provide a general framework derived by modifying Polyanskiy et. al. meta converse [7] and specifically applying it to EH-AWGN channels. Note that the converse is proved under the maximal probability of error criterion. We are able to show that in both, the achievability and converse, the second order term is  $O(\sqrt{n})$ . This also gives us the strong converse for this channel for free as the first order term is unaffected by the probability of error term. Finally, we analyze DMCs with energy harvesting and provide the finite blocklength achievability and converse bounds for them.

The paper is organized as follows. Then we provide the notation and some basic results in section II. We next give the details of the energy harvesting system for the AWGN channel in section III. Section IV deals with the finite blocklength achievability for an EH-AWGN channel and section V deals with the finite blocklength converse for an EH-AWGN channel. In section VI, we derive the achievability and converse bounds for an energy harvesting DMC. After that we conclude the paper.

## II. PRELIMINARIES

### A. Basic notation

We shall use boldface letters (e.g.  $\mathbf{x}$ ) to denote vectors (belonging to  $\mathbb{R}^n$  for a specified  $n \in \mathbb{N}$ ). When the length of the vector needs to be specified, we shall mention it as  $x^{(k)} = (x_1, x_2, \dots, x_k)$ . Lowercase letters denote deterministic scalars or vectors whereas uppercase letters denote random variables or random vectors respectively. We shall use  $[M]$  to denote the set  $\{1, 2, \dots, M\}$ . We shall denote  $\mathcal{P}(\mathcal{X})$  for the set of probability distributions on  $\mathcal{X}$  (In cases where the alphabet is clear, we simply use  $\mathcal{P}$ ). The expected operator will be denoted by  $\mathbb{E}$  and if the distribution (say  $P$ ) needs to be specified, then it shall be denoted as  $\mathbb{E}_P$ .

### B. Channels, probability of error and capacity

Given an input alphabet  $\mathcal{X}$  and output alphabet  $\mathcal{Y}$ , a *channel*, denoted by  $W(y|x)$ , is a conditional probability measure on  $\mathcal{Y}$  given  $x \in \mathcal{X}$ . If the probability density function exists for the channel (call it  $f_{Y|X}$ ), then for any measurable set  $B \subseteq \mathcal{Y}$ ,

$$W(B|x) = \int_B f_{Y|X}(y|x) dy = \int_B dW(y|x). \quad (1)$$

Given a probability distribution  $P$  on  $\mathcal{X}$  and a channel  $W$ , we define the measure  $PW$  as

$$PW(y) = \sum_{x \in \mathcal{X}} P(x)W(y|x) \quad (2)$$

There are two notions of probability of error which we will use. Given a code  $\mathcal{C}$  with  $M$  messages, let  $U \in [M]$  be the random variable denoting the message to be transmitted and  $\hat{U} \in [M]$  the message that was decoded at the receiver. Then given a channel  $W(\mathbf{y}|\mathbf{x})$ , the *maximum probability of error* of the code  $\mathcal{C}$  is defined as

$$P_{e,max}(\mathcal{C}) := \max_{1 \leq m \leq M} Pr [\hat{U} \neq m | U = m]. \quad (3)$$

It is convenient to work with the probability of correct decoding ( $1 - P_{e,max}$ ) as will be clear from the discussions that follow. Similarly the *average probability of error* is defined as

$$P_{e,avg}(\mathcal{C}) := \frac{1}{M} \sum_{m=1}^M Pr [\hat{U} \neq m | U = m]. \quad (4)$$

Clearly for any given code,  $P_{e,avg} \leq P_{e,max}$ . While these two notions are different, it can be shown that as far as channel capacity (defined below) is concerned, there is no effect. However, in the finite blocklength regime, there are effects but only in higher order terms. Hence we may pick any one criterion and analyze with respect to that.

An  $(n, M, \varepsilon)$  code is a code with  $M$  codewords of codeword length  $n$  and probability of error at most  $\varepsilon$ . We can define

$$M^*(n, \varepsilon) := \max\{M : \text{There exists a } (n, M, \varepsilon) \text{ code}\}. \quad (5)$$

Given a  $(n, M, \varepsilon)$  code, we shall call  $\frac{\log M}{n}$  as the *rate* of the code. For  $0 < \varepsilon < 1$ , the  $\varepsilon$ -*capacity*  $C_\varepsilon$  is defined as

$$C_\varepsilon = \lim_{n \rightarrow \infty} \frac{\log M^*(n, \varepsilon)}{n} \quad (6)$$

and the *capacity* of the channel is defined as

$$C = \lim_{\varepsilon \rightarrow 0} C_\varepsilon. \quad (7)$$

Note that both limits exist. It is clear that  $C_\varepsilon \geq C$ . However for certain classes of channels like DMCs and standard AWGN channels with average power constraints, we have  $C_\varepsilon = C$  for every  $0 < \varepsilon < 1$ . Then we say that the channel satisfies the *strong converse*, which means that if we transmit at rates greater than capacity, the probability of error of the code tends to 1 as the blocklength  $n$  tends to infinity.

### C. AWGN Channel

Given  $\mathbf{a} \in \mathbb{R}^n$  and a covariance matrix  $\mathbf{K} \in \mathbb{R}^{n \times n}$ , define

$$\mathcal{N}(\mathbf{a}; \mathbf{K}) := \frac{\exp \{ -(\mathbf{x} - \mathbf{a})^T \mathbf{K}^{-1} (\mathbf{x} - \mathbf{a}) \}}{(2\pi)^{n/2} (\det(\mathbf{K}))^{1/2}} \quad (8)$$

as the multivariate normal distribution with mean  $\mu$  and covariance matrix  $\mathbf{K}$  whose determinant is non-zero. An Additive White Gaussian Noise (AWGN) channel with noise variance  $\sigma^2$  is given by

$$Y = x + Z$$

where  $x \in \mathbb{R}$  is the input to the channel and  $Z \sim \mathcal{N}(0; \sigma^2)$ . The  $n$ -dimensional version is obtained by applying the one dimensional version ( $n = 1$ ) case independently on each input  $x_i$ ,  $1 \leq i \leq n$ . The AWGN channel with average power constraint  $S$  is an AWGN channel where the input  $\mathbf{x}$  satisfies

$$\|\mathbf{x}\|_2^2 \leq nS \quad (9)$$

where for  $p \geq 1$ ,  $\|\mathbf{x}\|_p = \left( \sum_{i=1}^n x_i^p \right)^{1/p}$  is the  $p$ th norm of  $\mathbf{x}$ .

The capacity of an AWGN channel (denoted by  $C_G$ ) with average power constraint  $P$  is given by

$$C_G := \frac{1}{2} \log_2 \left( 1 + \frac{P}{\sigma^2} \right) \text{ bits per channel use.} \quad (10)$$

Moreover it is known that the strong converse holds for standard AWGN channels and so the  $\varepsilon$ -capacity is equal to  $C_G$ . This is also evident from the finite blocklength characterization of capacity (see [5],[6]) where it was shown that for an AWGN channel with average power constraint  $P$ , the maximum code size  $M^*(n, \varepsilon, P)$ , for  $n$  sufficiently large, satisfies

$$\log M^*(n, \varepsilon, P) = nC_G + \sqrt{nV_G} \Phi^{-1}(\varepsilon) + O(\log(n)) \quad (11)$$

where the probability of error is at most  $\varepsilon$ ,  $V_G = \frac{P(P+2)}{2(P+1)^2}$  and  $\Phi(x) = \int_{-\infty}^x \frac{e^{-u^2/2}}{\sqrt{2\pi}} du$ .

### D. Discrete Memoryless Channels (DMC)

A DMC is characterized by a finite input alphabet  $\mathcal{X}$ , finite output alphabet  $\mathcal{Y}$  and the transition probabilities given by  $W = P_{Y|X}$ , which satisfies for every  $n \geq 1$

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n P_{Y|X}(y_i|x_i) \quad (12)$$

In short, the present output depends only on the present input. While the output is, in principle, allowed to depend on past outputs (which is known as a DMC with feedback), we only consider DMC's without feedback. The capacity  $C_D$  of a DMC  $W$  is given by Shannon's formula as

$$C_D = \sup_{P \in \mathcal{P}(\mathcal{X})} I(P; W) = \sup_{P \in \mathcal{P}(\mathcal{X})} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x) W(y|x) \log \left( \frac{W(y|x)}{PW(y)} \right), \quad (13)$$

where  $I(P; W)$  is the mutual information (see [14]) between the input and output of the channel.

It was shown by Wolfowitz that DMC's have a strong converse. This also follows from the finite blocklength characterization of capacity of DMC's given in [5] and later refined by [6], [7] and [8]. Before we recall the result, however, we define the information density of a channel (see [15] for more details) as follows.

**Definition 1.** Given a channel  $W$  and an output distribution  $Q$ , the information density of the channel is given by

$$i(x, y; Q) = \log \left( \frac{W(y|x)}{Q(y)} \right). \quad (14)$$

Often  $Q = PW$  for some  $P \in \mathcal{P}(\mathcal{X})$ , in which case we shall denote the information density by  $i_P(x, y)$ .

Observe that  $I(P; W) = \sum_{x,y} P(x) W(y|x) i_P(x, y)$ . Similarly, the variance of information density is given by

$$V(P; W) := \left[ \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x) W(y|x) (i_P(x, y))^2 \right] - (I(P; W))^2. \quad (15)$$

The finite blocklength result for DMC's with channel  $W$ , probability of error  $0 < \varepsilon < 1$  and  $V(P; W) > 0$  for some capacity achieving distribution  $P$  is given by

$$\log M^*(n, \varepsilon) = nC_D + \sqrt{nV_D}\Phi^{-1}(\varepsilon) + O(\log(n)), \quad (16)$$

where

$$V_D = \begin{cases} V_{\min} := \min_{P \in \Pi} V(P; W) & \varepsilon \leq 1/2 \\ V_{\max} := \max_{P \in \Pi} V(P; W) & \varepsilon > 1/2 \end{cases}. \quad (17)$$

and  $\Pi = \{P \in \mathcal{P}(\mathcal{X}) : I(P; W) = C_D\}$  is the set of capacity achieving distributions.

#### E. DMC with cost constraints

Let  $\Lambda : \mathcal{X} \rightarrow \mathbb{R}$  be a non-negative function which we will refer to as the energy function. The energy function simply returns the energy of the symbol  $x$  which is a generalization of the standard energy function  $\Lambda(x) = x^2$  for an AWGN channel. We further assume that the energy function is separable, i.e. given a vector  $\mathbf{x} \in \mathcal{X}^n$ ,

$$\Lambda(\mathbf{x}) := \sum_{i=1}^n \Lambda(x_i). \quad (18)$$

Define the constrained sets  $\mathbb{F}_a$  and  $\mathcal{F}_a$  for  $a \geq 0$  as follows

$$\mathbb{F}_a = \{\mathbf{x} \in \mathcal{X}^n : \Lambda(\mathbf{x}) \leq na\}, \quad (19)$$

$$\mathcal{F}_a = \{P \in \mathcal{P} : \mathbb{E}_P[\Lambda(X)] \leq a\}. \quad (20)$$

In a DMC with cost constraints(see [16], [17]), where the codewords are drawn from  $\mathbb{F}_a$ , the capacity changes to

$$C_F(a) = \sup_{P \in \mathcal{F}_a} I(P; W). \quad (21)$$

Moreover, the maximum achievable code size, for any  $a > 0$ , denoted by  $M^*(n, \varepsilon, a)$ , under some regularity conditions (see [6] for the result and [17] for some refinements), is given by

$$\log M^*(n, \varepsilon, a) = nC_F(a) + \sqrt{nV_F(a)}\Phi^{-1}(\varepsilon) + O(\log n) \quad (22)$$

where  $V_F(a)$  is the channel dispersion (see [6]).

An energy harvesting DMC (EH-DMC), may be viewed as a generalization of a DMC with cost constraints and its finite blocklength analysis is reserved to section VI.

### III. ENERGY HARVESTING AWGN CHANNEL

An energy harvesting system consists of an energy buffer which stores energy from various sources over a period of time. Energy is usually harvested from some ambient source, e.g., solar power. An EH-AWGN channel consists of an energy harvesting system at the transmitter end, followed by an AWGN channel as shown in Fig. 1. The idea here is that symbol

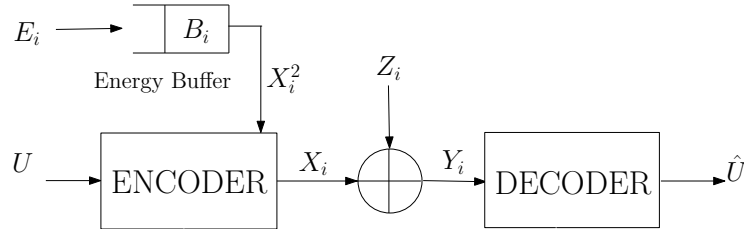


Fig. 1. Block diagram of an AWGN energy harvesting system

transmissions consume energy. For instance to send symbol  $x$  on the channel, we would require  $x^2$  units of energy from the buffer and if sufficient energy is available, the transmission succeeds; otherwise it fails. We assume that the energy buffer has infinite capacity, i.e., there is always sufficient space to store energy and energy leakages do not occur. We assume that the incoming energy process  $(E_i)$  is i.i.d., non negative with finite mean  $\mathbb{E}[E_1]$  and finite variance  $\sigma_E^2$ .

There are several energy harvesting paradigms for an energy harvesting channel. For example in [10], the Harvest-Use (HU), Harvest-Store-Use (HSU) and Harvest-Use-Store (HUS) paradigms are studied. In this paper, we focus on the HUS paradigm where in a transmission slot, we first harvest energy  $E_i$  in slot  $i$ , use it along with some energy in the buffer if needed to

transmit the symbol and then store the remaining energy. Let  $B_i$  be the energy in buffer at the  $i$ th transmission slot. Assume  $B_0 = 0$ . Then the energy in buffer, under HUS, for  $1 \leq i \leq n$  evolves as

$$B_i = (B_{i-1} + E_i - X_i^2)^+ \quad (23)$$

where  $(x)^+$  means  $\max\{x, 0\}$ . We note that a finite blocklength converse for the infinite buffer case will also be a converse for the finite buffer case. Essentially, the more constraints we have at the encoder end, lesser we expect the capacity to be. However, the bound is unlikely to be optimal for the finite buffer case.

For the ordinary AWGN channel with power constraint  $S$ , the sequences were supposed to satisfy (9). The constraint for the energy harvesting AWGN channel may be succinctly given as

$$\|x^{(k)}\|_2^2 \leq \|e^{(k)}\|_1 \quad 1 \leq k \leq n \quad (24)$$

where  $x^{(k)} = (x_1, x_2, \dots, x_k)$  is the first  $k$  codeword symbols and  $e^{(k)} = (e_1, e_2, \dots, e_k)$  represent the first  $k$  incoming energy symbols in the buffer. The constraint simply asserts the impossibility of transmitting a symbol when there is insufficient energy available, which we refer to as an *outage*. The capacity of an EH-AWGN channel (see [10] and [11]) is

$$C_{EG} = \frac{1}{2} \log \left( 1 + \frac{\mathbb{E}[E_1]}{\sigma^2} \right). \quad (25)$$

Additionally, the strong converse was also shown to hold for this channel (see [10]). This would logically imply a converse of the form  $\log M \leq nC_{EG} + o(n)$ . However as we seek a refinement of this expression, we would need finer tools to extract a finite blocklength converse. In this regard, we will be using several results from [7]. For clarity, we use the same notations in that paper.

#### IV. FINITE BLOCKLENGTH ACHIEVABILITY FOR EH-AWGN CHANNEL

We shall use the average probability of error criterion here. Let  $0 < \varepsilon < 1$  be given. We shall construct a code for the EH-AWGN channel, using random coding technique, which will have average probability of error not more than  $\varepsilon$ . Assume the buffer is empty at the beginning. The coding scheme we propose has two phases; namely a saving phase and a transmission phase. This is known in literature as the *save and transmit scheme* (see [11]).

##### A. Save and Transmit Scheme

In this scheme, the transmitter first transmits nothing (it is equivalent to transmitting the symbol 0) for a set number of slots. During this period, it allows the buffer to build up and gather energy. The receiver is aware of the number of slots and chooses to ignore the outputs since they are not information bearing. After this phase, the next phase is the transmission phase wherein symbols are transmitted as normal. Since we are starting with a buffer that isn't empty, this will improve the chances of non occurrence of energy outage. But the caveat is that slots are wasted in gathering energy. To ensure that this scheme does not affect the coefficient of the first order term, it is required that the number of slots set for gathering energy scale at most as  $o(n)$ .

We now state the EH-AWGN channel achievability theorem.

**Theorem 1.** *Given  $0 < \varepsilon < 1$ , under average probability of error, the maximal size of the code  $M^*(\hat{n}, \varepsilon)$  with blocklength  $\hat{n}$  for an EH-AWGN channel with variance  $\sigma^2$  with HUS architecture, and the energy process  $E_i$  i.i.d. with  $E[E_1^2] < \infty$ , satisfies the following bound*

$$\log M^*(\hat{n}, \varepsilon) \geq \hat{n}C_{EG} - \sqrt{\hat{n}}K_\varepsilon C_{EG} + \sqrt{\frac{\hat{n}V_{EG}}{2}}\Phi^{-1}(\lambda\varepsilon) - \log \hat{n} + O(1). \quad (26)$$

for any  $0 < \lambda < 1$ . Here  $C_{EG} = \frac{1}{2} \log_2 \left( 1 + \frac{\mathbb{E}[E_1]}{\sigma^2} \right)$ ,  $V_{EG} = \frac{E[E_1]}{E[E_1] + \sigma^2}$ ,  $K_\varepsilon = \frac{2\sqrt{\text{Var}(\Delta_1)}}{E[E_1]\sqrt{(1-\lambda)\varepsilon}}$  and  $\Delta_1 = E_1 - X_1^2$ .

The rest of this section is dedicated to proving this theorem.

##### B. Codebook Generation and Encoding

Let  $M$  be the number of messages which are assumed uniformly distributed. We will eventually derive a lower bound on  $M$ . Generate a matrix of size  $M \times n$  where each entry is generated i.i.d. with density  $\mathcal{N}(0; \mathbb{E}[E_1])$ . Denote each row by  $\mathbf{X}(m) = X^{(n)}(m)$ ,  $1 \leq m \leq M$ . This codebook is available at the decoder.

Let  $N_n$  denote the number of slots in the energy gathering phase as described earlier. During this phase, the buffer fills up with energy and after  $N_n$  time slots, we expect it to have crossed some threshold energy value which we will denote by  $E_{0n}$ . Let  $N_n = K_\varepsilon \sqrt{n}$ , where  $K_\varepsilon$  will be chosen later, and  $E_{0n} = N_n \mathbb{E}[E_1]/2$ .

Now after the gathering phase we have the transmission phase. Let  $n$  be the number of slots wherein we transmit symbols on the AWGN channel. We count channel uses from the  $N_n + 1$  instant onwards. Once we gather at least  $E_{0n}$  energy, we must ensure that subsequent transmissions will not cause an outage. Let us denote the energy constraints by  $\mathcal{A}_k$  where

$$\mathcal{A}_k = \bigcap_{l=1}^k \{S_l \geq -E_{0n}\}.$$

where  $S_l = \sum_{k=1}^l E_k - X_k^2$ . To send message  $m$ , at channel use  $k$ , where  $1 \leq k \leq n$ , we transmit  $\hat{X}_k(m) = X_k(m)1_{\mathcal{A}_k}$ . Note that the transmitted codeword satisfies the energy harvesting conditions, since  $E_{0n}$  energy has already been harvested before the transmission started. At the encoder end, if the energy gathered is less than  $E_{0n}$ , we declare an error. An error is also declared during transmission phase if there is insufficient energy to transmit a symbol.

### C. Decoder Design

The decoder ignores the output of the channel for the first  $N_n$  channel uses as those are for harvesting energy. During the transmission stage, i.e. from the  $(N_n + 1)$ st stage onwards, it retrieves the output of the channel  $\mathbf{W} = W^{(n)}$  and then decodes the unique message  $\hat{M} = m$  such that

$$\frac{1}{n} \log \left( \frac{\hat{\mathbf{W}}^n(\mathbf{W}|\mathbf{X}(m))}{\overline{\mathbf{W}}^n(\mathbf{W})} \right) > \frac{1}{n} \log(M) + \eta_n, \quad (27)$$

where  $\hat{\mathbf{W}}(\cdot|\mathbf{x}) = \mathcal{N}(\mathbf{x}; \sigma^2 \mathbf{I}_n)$  is the channel's transition probability and  $\overline{\mathbf{W}}^n(\cdot) = \mathcal{N}(0; (\mathbb{E}[E_1] + \sigma^2) \mathbf{I}_n)$ . We choose  $\eta_n > 0$  later. If such a message can't be found, the decoder randomly picks one of  $M$  possible messages. This decoder is known as the Threshold Decoder [13] (also see [7]).

### D. List of Error Events

Denote by  $\mathcal{E}$  the event that an error happens. Due to symmetry in the codebook construction, it suffices to assume that message 1 is sent. Let  $U$  be the random variable which is uniformly distributed indicating the message to be sent and  $\hat{U}$  the decoded message. Thus,

$$Pr(\mathcal{E}) = Pr(\hat{U} \neq 1 | U = 1).$$

where  $Pr(A)$  denotes the probability of event  $A$ .

Let,

$$\mathcal{E}_0 = \{S_{N_n}^E < E_{0n}\}, \quad (28)$$

$$\mathcal{E}_1 = \bigcup_{k=1}^n \{S_k < -E_{0n}\}, \quad (29)$$

$$\mathcal{E}_2 = \bigcup_{m=2}^M \left\{ \frac{1}{n} i(\mathbf{X}(m); \mathbf{W}) > \frac{1}{n} \log(M) + \eta_n \right\}, \quad (30)$$

$$\mathcal{E}_3 = \left\{ \frac{1}{n} i(\mathbf{X}(1); \mathbf{W}) \leq \frac{1}{n} \log(M) + \eta_n \right\}, \quad (31)$$

where  $S_{N_n}^E = \sum_{i=1}^{N_n} E_i$  and  $i(\mathbf{x}; \mathbf{w}) = \log \left( \frac{\hat{\mathbf{W}}^n(\mathbf{w}|\mathbf{x})}{\overline{\mathbf{W}}^n(\mathbf{w})} \right)$ . Here  $\mathcal{E}_0$  is the event that energy less than  $E_{0n}$  was harvested in the gathering phase;  $\mathcal{E}_1$  is the event that an energy outage occurred while transmitting the message;  $\mathcal{E}_2$  is the event that the decoder declares some other message  $m \neq 1$  as the sent message and finally  $\mathcal{E}_3$  is the event that the decoder failed to declare the first message as the one sent. Hence we have

$$Pr(\mathcal{E}) \leq \sum_{i=0}^3 Pr(\mathcal{E}_i). \quad (32)$$

Before we proceed with the error analysis, we give a few remarks.

- The input to the channel is  $\hat{X}_k(m) = X_k(m)1_{\mathcal{A}_k}$ . While  $X_k(m)$  is i.i.d. for  $1 \leq k \leq n$ ,  $\hat{X}_k(m)$  is not. This affects the analysis of  $Pr(\mathcal{E}_3)$ . However  $Pr(\mathcal{E})$  may be decomposed as

$$Pr(\mathcal{E}) \leq Pr(\mathcal{E}_0) + Pr(\mathcal{E}_1) + Pr(\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_0^c) + Pr(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_0^c)$$

Now  $\mathcal{E}_0^c \cap \mathcal{E}_1^c$  is precisely the event of no outage. Under this event,  $\hat{X}_k(m) = X_k(m)$ . Thus we may replace  $\hat{X}_k(m)$  with  $X_k(m)$  in events  $\mathcal{E}_2$  and  $\mathcal{E}_3$  and using the upper bound  $Pr(\mathcal{E}_i \cap \mathcal{E}_1^c \cap \mathcal{E}_0^c) \leq Pr(\mathcal{E}_i)$  for  $i = 2, 3$  after doing the replacement makes it easier to analyze  $Pr(\mathcal{E}_2)$  and  $Pr(\mathcal{E}_3)$ .

- The term  $i(\mathbf{x}; \mathbf{w})$  is called as the information density of the channel, which is basically the log likelihood ratio between the channel probability and an independent test channel. More details on this may be found in [15]. It is part of the threshold detector which by itself is motivated by general achievability lemmas like Feinstein's Lemma and Shannon's Lemma.
- The key difference between the approach mentioned here and in [13] is that in the latter, the probability of events  $\mathcal{E}_0$  and  $\mathcal{E}_1$  are combined together and bounded. However, in our case, we split it into two events and bound them individually. The individual bounds may be tighter, as is true in our case, as opposed to bounding together.

#### E. Error Event $\mathcal{E}_0$

From Chebyshev's inequality, noting that  $\sigma_E^2 < \infty$ , we have

$$\begin{aligned}
 Pr(\mathcal{E}_0) &= Pr[S_{N_n}^E < E_{0n}] \\
 &= Pr[S_{N_n}^E - N_n \mathbb{E}[E_1] < E_{0n} - N_n \mathbb{E}[E_1]] \\
 &\leq Pr[|S_{N_n}^E - N_n \mathbb{E}[E_1]| \geq N_n \mathbb{E}[E_1] - E_{0n}] \\
 &\leq \frac{N_n \sigma_E^2}{(N_n \mathbb{E}[E_1] - E_{0n})^2} \\
 &= \frac{4\sigma_E^2}{K_\varepsilon \sqrt{n} (\mathbb{E}[E_1])^2}
 \end{aligned} \tag{33}$$

#### F. Error Event $\mathcal{E}_1$

Let  $X_k = X_k(1)$  and  $S_n = \sum_{k=1}^n \Delta_k$ , where  $\Delta_k = E_i - X_i^2$ . Hence we have,

$$Pr(\mathcal{E}_1) = Pr\left[\bigcup_{k=1}^n \{S_k < -E_{0n}\}\right] \tag{34}$$

$$= Pr\left[\min_{k=1,2,\dots,n} S_k < -E_{0n}\right]. \tag{35}$$

Note that  $\Delta_k$  are i.i.d. and  $\mathbb{E}[\Delta_k] = \mathbb{E}[E_k] - \mathbb{E}[X_k^2] = 0$ . This means  $S_n$  is a zero drift random walk. Thus, by Kolmogorov's Inequality (see Chapter 3 of [18]), for any  $v > 0$ , we have

$$Pr\left(\min_{k=1,2,\dots,n} S_k < -v\right) = Pr\left(\max_{k=1,2,\dots,n} -S_k > v\right) \tag{36}$$

$$\leq Pr\left(\max_{k=1,2,\dots,n} |S_k| > v\right) \tag{37}$$

$$\leq \frac{Var(S_n)}{v^2} \tag{38}$$

$$= \frac{nVar(\Delta_1)}{v^2}. \tag{39}$$

Note that for the RHS to be a useful bound, it is required that  $v$  be at least  $O(\sqrt{n})$ . Since  $Var(\Delta_1) < \infty$ , picking  $v = E_{0n}$ , we get

$$Pr(\mathcal{E}_1) \leq \frac{nVar(\Delta_1)}{E_{0n}^2} = \frac{4Var(\Delta_1)}{K_\varepsilon^2 (\mathbb{E}[E_1])^2}. \tag{40}$$

While the RHS is not necessarily less than 1, by picking  $K_\varepsilon$  large enough and independent of  $n$ , we can suitably bound this error term. Fix  $0 < \lambda < 1$ . We shall choose  $K_\varepsilon$  as follows

$$K_\varepsilon = \frac{2\sqrt{Var(\Delta_1)}}{\mathbb{E}[E_1]\sqrt{(1-\lambda)\varepsilon}} \tag{41}$$

This will ensure that  $Pr(\mathcal{E}_1) \leq (1-\lambda)\varepsilon$ . An interesting aspect of bounding  $\mathcal{E}_1$  in this way is that this bound is independent of input distribution and depends only on the first two moments. Thus this method is useful when dealing with channels where the achievability proof starts with a input distribution which may not be tractable.

### G. Error Event $\mathcal{E}_2$

The error probability  $Pr(\mathcal{E}_2)$  may be upper bounded by a lemma proved by Shannon in [19] (also see [15]). We provide the proof here for completeness. Let

$$\mathcal{D}_{m,n} = \left\{ \frac{1}{n} i(\mathbf{X}(m); \mathbf{W}) > \frac{1}{n} \log(M) + \eta_n \right\}.$$

Then

$$\begin{aligned} Pr(\mathcal{E}_2) &\leq Pr\left(\bigcup_{m=2}^M \mathcal{D}_{m,n}\right) \\ &\leq \sum_{m=2}^M Pr(\mathcal{D}_{m,n}) \\ &= \sum_{m=2}^M \int f_{X^n}(\mathbf{x}) \bar{\mathbf{W}}^n(\mathbf{w}) 1_{\mathcal{D}_{m,n}} d\mathbf{x} d\mathbf{w} \\ &\leq \sum_{m=2}^M \int f_{X^n}(\mathbf{x}) \hat{\mathbf{W}}^n(\mathbf{w}|\mathbf{x}) \frac{2^{-n\eta_n}}{M} d\mathbf{x} d\mathbf{w} \\ &\leq 2^{-n\eta_n} \end{aligned} \tag{42}$$

where  $f_{X^n}(\mathbf{x}) = \prod_{i=1}^n f_X(x_i)$ .

### H. Error Event $\mathcal{E}_3$

Let  $G_i = \log\left(\frac{\mathbf{W}^1(W_i|X_i(1))}{\bar{\mathbf{W}}^1(W_i)}\right)$ . Then we have

$$Pr(\mathcal{E}_3) \leq Pr\left\{\sum_{i=1}^n G_i \leq \log(M) + n\eta_n\right\}. \tag{43}$$

Note that  $G_i$  are i.i.d based on the remarks provided earlier. Moreover, we have

$$C_{EG} := E[G_i] = \frac{1}{2} \log\left(1 + \frac{\mathbb{E}[E_1]}{\sigma^2}\right), \tag{44}$$

$$V_{EG} := Var(G_i) = \frac{\mathbb{E}[E_1]}{\mathbb{E}[E_1] + \sigma^2}. \tag{45}$$

Also the third moment,  $E[|G_i|^3]$ , is finite. To proceed further, we state the Berry Esseen's theorem (see Theorem 6.4.1 in [18]).

**Lemma 1** (Berry Esseen's Theorem). *Let  $X_i$ ,  $1 \leq i \leq n$ , be an i.i.d. sequence of random variables with mean  $\mu$ , variance  $\sigma^2 < \infty$  and  $E[|X_1|^3] < \infty$ . Let  $S_n = \sum_{i=1}^n X_i$ . Then we have, for any  $x \in \mathbb{R}$ ,*

$$\left| Pr\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) - \Phi(x) \right| \leq C \frac{E|X_1 - \mu|^3}{\sigma^3\sqrt{n}},$$

where  $C < 1/2$  (see [20]). Note that the bound is uniform in  $x$ .

Let  $K = \frac{E[|G_i - E[G_i]|^3]}{2V_{EG}^{3/2}}$ . Applying Berry Esseen's theorem, we have for any  $u \in \mathbb{R}$ ,

$$\left| Pr\left\{\frac{\left(\sum_{i=1}^n G_i\right) - nC_{EG}}{\sqrt{nV_{EG}}} \leq u\right\} - \Phi(u) \right| \leq \frac{K}{\sqrt{n}}.$$

Substituting  $u = \frac{\log M + n(\eta_n - C_{EG})}{\sqrt{nV_{EG}}}$ , we get

$$Pr\left\{\sum_{i=1}^n G_i \leq \log(M) + n\eta_n\right\} \leq \Phi\left(\frac{\log M + n(\eta_n - C_{EG})}{\sqrt{nV_{EG}}}\right) + \frac{K}{\sqrt{n}}. \tag{46}$$

Let

$$\varepsilon_n = \lambda\varepsilon - \frac{4\sigma_E^2}{\sqrt{n}(K_\varepsilon\mathbb{E}[E_1])^2} - 2^{-n\eta_n} - \frac{K}{\sqrt{n}}. \tag{47}$$



By picking  $n$  large enough, we can ensure  $\varepsilon_n > 0$ .

Now if we determine conditions that will ensure  $Pr(\mathcal{E}_3) \leq \varepsilon_n$ , it will imply that  $Pr(\mathcal{E}) < \varepsilon$ , which is what we require. Taking

$$\log M \leq nC_{EG} + \sqrt{nV_{EG}}\Phi^{-1}(\varepsilon_n) - n\eta_n \quad (48)$$

for  $n$  large enough, will ensure that the RHS of (46) is less than  $\varepsilon_n + \frac{K}{\sqrt{n}}$ . This indicates that as long as  $M$  satisfies the above equation, the average probability of error can be made less than or equal to  $\varepsilon$ . Since the optimum code size can only be larger than the RHS of (48), we conclude that the largest code size satisfies

$$\log M^*(n + N_n, \varepsilon) \geq nC_{EG} + \sqrt{nV_{EG}}\Phi^{-1}(\varepsilon_n) - n\eta_n - 1. \quad (49)$$

We further simplify  $\Phi^{-1}(\varepsilon_n)$  using Taylor's theorem. There exists  $u \in (\varepsilon_n, \lambda\varepsilon)$  such that

$$f(\varepsilon_n) = f(\lambda\varepsilon) + (\varepsilon_n - \lambda\varepsilon)f'(u),$$

where  $f(x) = \Phi^{-1}(x)$ . Note that  $f(x)$  has a derivative that is positive, strictly decreasing upto  $x = 1/2$ ; beyond which it increases. Thus in  $(\varepsilon_n, \varepsilon)$ ,  $f'(u) \leq \hat{f} = \max\{f'(\varepsilon_{n_0}), f'(\varepsilon)\}$  where  $n_0$  is the smallest  $n$  for which  $\varepsilon_n > 0$ . Hence we get

$$\log M^*(n + N_n, \varepsilon) \geq n(C_{EG} - \eta_n) + \sqrt{nV_{EG}}\Phi^{-1}(\lambda\varepsilon) + \sqrt{nV_{EG}}(\varepsilon_n - \lambda\varepsilon)\hat{f} - 1.$$

We pick  $\eta_n = \frac{\log(n)}{n}$ . Hence we observe,

$$\sqrt{n}(\varepsilon_n - \lambda\varepsilon) = -\frac{4\sigma_E^2}{K_\varepsilon^2(\mathbb{E}[E_1])^2} - \frac{1}{\sqrt{n}} - K. \quad (50)$$

Gathering all terms together, we get

$$\log M^*(n + N_n, \varepsilon) \geq nC_{EG} + \sqrt{nV_{EG}}\Phi^{-1}(\lambda\varepsilon) - \log n + O(1). \quad (51)$$

Substituting  $\hat{n} = n + N_n$ ,

$$\begin{aligned} \log M^*(\hat{n}, \varepsilon) &\geq (\hat{n} - N_n)C_{EG} + \sqrt{(\hat{n} - N_n)V_{EG}}\Phi^{-1}(\lambda\varepsilon) - \log(\hat{n} - N_n) + O(1) \\ &\geq \hat{n}C_{EG} - N_nC_{EG} + \sqrt{(\hat{n} - N_n)V_{EG}}\Phi^{-1}(\lambda\varepsilon) - \log(\hat{n}) + O(1). \end{aligned}$$

We note that for  $n$  large enough,  $\frac{N_n}{\hat{n}} < \frac{1}{2}$  and hence,

$$\sqrt{(\hat{n} - N_n)V_{EG}}\Phi^{-1}(\lambda\varepsilon) \geq \begin{cases} \sqrt{\hat{n}V_{EG}}\Phi^{-1}(\lambda\varepsilon) & 0 < \varepsilon < \frac{1}{2\lambda}, \\ \sqrt{\frac{\hat{n}}{2}V_{EG}}\Phi^{-1}(\lambda\varepsilon) & \frac{1}{2\lambda} < \varepsilon < 1, \end{cases}$$

Thus we get for every  $0 < \varepsilon < 1$ , and  $\hat{n}$  large enough

$$\log M^*(\hat{n}, \varepsilon) \geq \hat{n}C_{EG} - \sqrt{\hat{n}}K_\varepsilon C_{EG} + \sqrt{\frac{\hat{n}V_{EG}}{2}}\Phi^{-1}(\lambda\varepsilon) - \log \hat{n} + O(1) \quad (52)$$

for some constant  $\hat{C}$  which concludes this proof. Moreover dividing by  $\hat{n}$  gives us

$$\frac{\log M^*(\hat{n}, \varepsilon)}{\hat{n}} \geq C_{EG} - O\left(\frac{1}{\sqrt{\hat{n}}}\right).$$

Thus, we see that the backoff from capacity is  $O\left(\frac{1}{\sqrt{\hat{n}}}\right)$ . Note that we cannot take  $\lambda \rightarrow 1$  as that would imply  $K_\varepsilon \rightarrow \infty$ . But for best results, we can pick  $\lambda$  that gives the best lower bound. ■

## V. FINITE BLOCKLENGTH CONVERSE FOR ENERGY HARVESTING AWGN CHANNELS

In any converse proof, the first step is to start with the probability of error expression and eventually obtain a bound on maximum code size. Throughout our discussion, we shall consider the maximum probability of error criterion. The energy arrivals  $E_i$  are assumed i.i.d. with mean  $\mathbb{E}[E_1]$  and variance  $\sigma_E^2$  as before.

### A. Encoder and Decoder

Let  $0 < \varepsilon < 1$  be given. Let  $M$  be the number of messages we can transmit while incurring a maximum probability of error  $\varepsilon$ . We denote  $[M]$  as shorthand for  $\{1, 2, \dots, M\}$ . Let  $U \in [M]$  be a random variable denoting the message to be transmitted and let  $\hat{U}$  denote the message decoded at the decoder. We additionally assume that the messages are equiprobable. The codeword corresponding to the message  $m$ , denoted by  $c(m, \mathbf{e}) \in \mathcal{X}^n$ , is a function of  $\mathbf{e}$ , the incoming energy. The codewords are chosen so as to satisfy the energy harvesting constraints i.e. in  $n$  transmissions, we have

$$\|c(m, e^{(k)})\|_2^2 \leq \|e^{(k)}\|_1 \quad 1 \leq k \leq n. \quad (53)$$

We note that the following Markov chain holds.

$$(U, E) \leftrightarrow X \leftrightarrow Y \leftrightarrow \hat{U}, \quad (54)$$

noting that  $U$  is independent of  $E$ . As is the case with any converse proof, we start with a code with maximal probability of error  $\varepsilon$  that satisfies the system constraints and proceed to find an upper bound on  $M$ . We will be using methods from binary hypothesis testing in the following sections.

### B. $\beta$ error function

We are going to use an important error function (we call it the  $\beta$  error function) between two distributions. See [14] for more details although the notations are from [7].

**Definition 2.** Given two distributions  $P$  and  $Q$  on  $\mathcal{X}$ , define for  $\alpha \in [0, 1]$ ,

$$\beta_\alpha(P, Q) := \min Q[T = 1] := \min \int_{\mathcal{X}} P_{T|X}(1|x) dQ(x) \quad (55)$$

where the minimum is over all distributions  $(P_{T|X})$  of test functions  $T : \mathcal{X} \rightarrow \{0, 1\}$  such that  $P[T = 1] \geq \alpha$ .

This function is essentially the type 2 error probability (probability of deciding  $P$  when  $Q$  is true) when the type 1 error probability is less than  $1 - \alpha$ . In fact, if  $P^n$  and  $Q^n$  are the  $n$  fold independent product distributions of  $P$  and  $Q$  respectively, then for  $0 < \alpha < 1$ ,

$$\lim_{n \rightarrow \infty} \frac{-\log \beta_\alpha(P^n, Q^n)}{n} = D(P||Q). \quad (56)$$

The right hand side is simply the Kullback Liebler divergence between two distributions and this result is called Stein's lemma (see [14]). The following lemma provides a useful lower bound for the  $\beta$  error function.

**Lemma 2.** Given two distributions  $P$  and  $Q$  on  $\mathcal{X}$  such that  $P << Q$ , for any  $0 < \alpha < 1$ , it holds that

$$\beta_\alpha(P, Q) \geq \sup_{\gamma > 0} \left\{ \frac{\alpha - P\left[\frac{dP}{dQ} \geq \gamma\right]}{\gamma} \right\}. \quad (57)$$

*Proof of Lemma 2:* See [7]. ■

The  $\beta$  error function is used in the meta-converse which is discussed next.

### C. Meta Converse

The Meta Converse, proved in [7], is one of the tightest known general converse bounds for any channel. It is so named as it recovers many known tight converse bounds including the Wolfowitz converse and the Han-Verdu Lemma as special cases.

**Lemma 3 (Meta Converse).** Every  $(M, \varepsilon)$  maximal probability of error code satisfies the following bound for any output distribution  $Q_Y$  and codewords coming from  $\mathbb{F} \subset \mathcal{X}$ , where  $\mathcal{X}$  is the input alphabet.

$$M \leq \frac{1}{\beta_{1-\varepsilon}(P_{Y|X=c_{\bar{m}}}, Q_Y)} \leq \sup_{x \in \mathbb{F}} \frac{1}{\beta_{1-\varepsilon}(P_{Y|X=x}, Q_Y)}. \quad (58)$$

where

$$\bar{m} = \arg \min_{m \in [M]} \Pr[\hat{U} = m | U = m].$$

This form is preferred when the channel has constraints such as an average power constraint etc. When the constraint set is deterministic, e.g.  $\mathbb{F} = \{x^n : \|x^n\| \leq nP\}$ , one can directly apply (57) and (58), along with some probabilistic bounds like Berry Esseen's theorem to get the desired converse result. However in the energy harvesting problem, the constraint set is stochastic, because the energy harvesting process is random, and so we cannot directly use the results. Instead, we need to refine the Meta-Converse, for which we require the following two lemmas.

**Lemma 4.** Given  $E_i$  i.i.d. random variables with mean  $\mu_E$ , variance  $\sigma_E^2 < \infty$ , we have given any  $\delta_n > 0$ ,

$$\Pr \left[ \sum_{i=1}^n E_i \geq n(\mu_E + \delta_n) \right] \leq \frac{\sigma_E^2}{n\delta_n^2}. \quad (59)$$

The proof follows from Chebyshev's inequality. The next lemma gives an upper bound on the complementary cdf. of the information density between two Gaussian distributions.

**Lemma 5.** Let  $P_{Y^n|X^n=x^n} = \mathcal{N}(x^n; \sigma^2 I_n)$  and  $Q_{\mathbf{Y}} = \mathcal{N}(0; (S + \sigma^2)I_n)$  where  $S > 0$ ,  $I_n$  is the identity matrix of dimension  $n \times n$  and  $x^n \in \mathbb{F} := \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|_2^2 \leq nS\}$ . Then for any  $\zeta_n < 0$ ,

$$P_{Y^n|X^n=x^n} \left( \frac{dP_{Y^n|X^n=x^n}}{dQ_{\mathbf{Y}}} \geq 2^{nC_S - \zeta_n} \right) \leq \exp \left( -\frac{\zeta_n^2}{2nV_S} \right), \quad (60)$$

where  $C_S = \frac{1}{2} \log_2(1 + \frac{S}{\sigma^2})$  and  $V_S = \frac{S(S+2\sigma^2)}{2(S+\sigma^2)^2} \log_2^2(e)$ .

*Proof of Lemma 5:* Let  $\{Z_i\}$  be i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$ . Then we have for any  $\gamma_n > 0$ ,

$$P_{Y^n|X^n=x^n} \left( \frac{dP_{Y^n|X^n=x^n}}{dQ_{\mathbf{Y}}} \geq \gamma_n \right) \quad (61)$$

$$\begin{aligned} &= P_{Z^n} \left( nC_S + \left( \frac{\|x^n + Z^n\|_2^2}{2(S + \sigma^2)} - \frac{\|Z^n\|_2^2}{2\sigma^2} \right) \log_2(e) \geq \log \gamma_n \right) \\ &\leq P_{Z^n} \left( nC_S + \frac{S \log_2(e)}{2(S + \sigma^2)\sigma^2} \sum_{i=1}^n \left\{ -Z_i^2 + \frac{2Z_i x_i \sigma^2}{S} + \sigma^2 \right\} \geq \log \gamma_n \right) \end{aligned} \quad (62)$$

where in (62), we used that  $\|x^n\|_2^2 \leq nS$ . Let us set  $\rho = \frac{S \log_2(e)}{2(S + \sigma^2)}$ . Completing the squares and picking  $\gamma_n = 2^{nC_S - \zeta_n}$  for any  $\zeta_n \in \mathbb{R}$  yields the following expression.

$$\begin{aligned} &P_{Y^n|X^n=x^n} \left( \frac{dP_{Y^n|X^n=x^n}}{dQ_{\mathbf{Y}}} \geq \gamma_n \right) \\ &\leq \Pr \left( \sum_{i=1}^n \left( \frac{Z_i}{\sigma} - \frac{x_i \sigma}{S} \right)^2 \leq \frac{\zeta_n}{\rho} + n + \sum_{i=1}^n \frac{x_i^2 \sigma^2}{S^2} \right) \end{aligned} \quad (63)$$

Let  $\lambda_i = \left( \frac{Z_i}{\sigma} - \frac{x_i \sigma}{S} \right)^2$ . Then  $\sum_{i=1}^n \lambda_i$  is a non-central chi squared random variable with  $n$  degrees of freedom and noncentrality parameter  $\sum_{i=1}^n \frac{x_i^2 \sigma^2}{S^2}$ . We recall the following bound for non-central chi squared random variables ([21], Lemma 8.1).

**Lemma 6.** If  $\chi$  is a non-central chi-squared random variable with  $n$  degrees of freedom and non-centrality parameter  $\bar{B}$ , then for any  $t > 0$ ,

$$\Pr \left( \chi \leq n + \bar{B} - 2\sqrt{(n + 2\bar{B})t} \right) \leq e^{-t} \quad (64)$$

This lemma is actually the Chernoff bound optimized for non central chi squared random variables. Using this in (63) yields

$$\begin{aligned} &P_{Y^n|X^n=x^n} \left( \frac{dP_{Y^n|X^n=x^n}}{dQ_{\mathbf{Y}}} \geq \gamma_n \right) \\ &\leq \exp \left( -\frac{\zeta_n^2}{4\rho^2 \left( n + 2 \sum_{i=1}^n \frac{x_i^2 \sigma^2}{S^2} \right)} \right) \\ &\leq \exp \left( -\frac{\zeta_n^2}{4\rho^2 n \left( 1 + \frac{2\sigma^2}{S} \right)} \right) \end{aligned} \quad (65)$$

Note that here  $\zeta_n < 0$  which follows by comparing (63) and (64). Simplifying and substituting for  $V_S$  gives us the desired form.  $\blacksquare$

We now proceed to describe the main theorem of our paper.

**Theorem 2.** Given an energy harvesting AWGN channel with noise variance  $\sigma^2$ , energy arrival process  $E_i$  being i.i.d., non negative with mean  $E[E_1] < \infty$  and  $\text{Var}(E_1) = \sigma_E^2 < \infty$ , given maximal probability of error  $0 < \varepsilon < 1$ , the maximum code size is bounded as

$$\log M^* \leq nC_{EG} + \frac{\sqrt{n}D_\varepsilon \log_2(e)}{2(E[E_1] + \sigma^2)} + \sqrt{nV_{EG} \log \left( \frac{1}{(1 - \varepsilon)^2} \right)} + \sqrt{n(1 - \varepsilon)} + O(n^{1/4}). \quad (66)$$

where  $D_\varepsilon = \sqrt{\frac{4\sigma_E^2}{1-\varepsilon}}$ .

Before we prove the theorem for EH-AWGN channels, we will develop a general framework for energy harvesting converses. This is done in the following section.

#### D. The energy harvesting meta converse

The first step is to modify the meta converse appropriately. We first weaken the assumption on feasible codewords and assume that given blocklength  $n$ ,

$$\|c(m, e^n)\|_2^2 \leq \|e^n\|_1. \quad (67)$$

Note that the maximum code size can only increase if we do this and so, any upper bound on  $M$  for this problem will be an upper bound for the original problem.

Under the aforementioned condition, we modify the meta-converse appropriately and state the result as the following lemma.

**Lemma 7.** *For an energy harvesting channel, given an  $(M, \varepsilon)$  code for  $0 < \varepsilon < 1$ , the energy harvesting process having first three moments finite, we have for any distribution  $Q_{\mathbf{Y}}$  and  $\delta_n > 0$ ,*

$$M \leq \frac{1}{\inf_{x^n \in \mathbb{F}} \beta_{1-\varepsilon-\tau_n} (P_{Y^n|X^n}(\cdot|x^n), Q_{\mathbf{Y}})}. \quad (68)$$

where  $\mathbb{F} = \{x^n : \|x^n\|_2^2 \leq n(\mathbb{E}[E_1] + \delta_n)\}$ ,  $\mathcal{E}_1 = \{e^n : \|e^n\| \leq n(\mathbb{E}[E_1] + \delta_n)\}$  and  $\tau_n = P_{E^n}(\mathcal{E}_1^c)$ .

*Proof of Lemma 7:* The proof follows the steps used in proving the original meta-converse (see [7]) upto a point. Given distribution  $Q_{\mathbf{Y}}$ , which is essentially a reference channel that does not depend on input, let the maximal probability of error for this “channel” be  $\varepsilon'$ . Let  $U$  be the random variable denoting the message to be sent and  $\hat{U}$  be the message that was decoded.

Consider the definition of maximal probability of error. We see that there is a message, call it  $\bar{m}$  such that

$$1 - \varepsilon' = \Pr \left[ \hat{U} = \bar{m} | U = \bar{m} \right] = \int_y P_{\hat{U}|Y^n}(\bar{m}|y^n) dQ_{\mathbf{Y}}(y^n). \quad (69)$$

But we also have

$$1 - \varepsilon' = \min_m \Pr \left[ \hat{U} = m | U = m \right] \quad (70)$$

$$\leq \frac{1}{M} \sum_{m=1}^M \Pr \left[ \hat{U} = m | U = m \right] \quad (71)$$

$$= \frac{1}{M} \sum_{m=1}^M \int_{y^n} \Pr \left[ \hat{U} = m | Y^n = y^n \right] dQ_{\mathbf{Y}}(y^n) \quad (72)$$

$$= \frac{1}{M} \int_{y^n} \left( \sum_{m=1}^M \Pr \left[ \hat{U} = m | Y^n = y^n \right] \right) dQ_{\mathbf{Y}}(y^n) \quad (73)$$

$$= \frac{1}{M} \quad (74)$$

Combining equation (69) and (74), we get

$$M \leq \frac{1}{\int_y P_{\hat{U}|Y^n}(\bar{m}|y^n) dQ_{\mathbf{Y}}(y^n)}. \quad (75)$$

Now we have for any  $\mathcal{E}_1 \subset \mathbb{R}_+^n$ ,

$$1 - \varepsilon \leq \int_{e^n} \int_{y^n} P_{\hat{U}|Y^n}(\bar{m}|y^n) dP_{Y^n|X^n}(y^n|c(\bar{m}, e^n)) dP_E^n(e^n) \quad (76)$$

$$\leq \int_{e^n \in \mathcal{E}_1} \int_{y^n} P_{\hat{U}|Y^n}(\bar{m}|y^n) dP_{Y^n|X^n}(y^n|c(\bar{m}, e^n)) dP_E(e^n) + P_E(\mathcal{E}_1^c). \quad (77)$$

Rearranging and using the definitions given in the statement of the lemma, we get

$$1 - \varepsilon - \tau_n \leq \int_{e^n \in \mathcal{E}_1} \int_{y^n} P_{\hat{U}|Y^n}(\bar{m}|y^n) dP_{Y^n|X^n}(y^n|c(\bar{m}, e^n)) dP_{E^n}(e^n) \quad (78)$$

$$\Rightarrow 1 - \varepsilon - \tau_n \leq \frac{1 - \varepsilon - \tau_n}{1 - \tau_n} \leq \int_{y^n} P_{\hat{U}|Y^n}(\bar{m}|y^n) \left\{ \int_{e^n \in \mathcal{E}_1} dP_{Y^n|X^n}(y^n|c(\bar{m}, e^n)) \frac{dP_{E^n}(e^n)}{1 - \tau_n} \right\}. \quad (79)$$

Note that we divide by  $1 - \tau_n$  is to ensure that the term in braces is a probability distribution. From (75), (79) and the definition of  $\beta$  error function, we get

$$\frac{1}{M} \geq \beta_{1-\varepsilon-\tau_n} \left( \int_{e^n \in \mathcal{E}_1} dP_{Y^n|X^n}(\cdot|c(\bar{m}, e^n)) \frac{dP_{E^n}(e^n)}{1 - \tau_n}, Q_Y \right) \quad (80)$$

$$\geq \inf_{x^n \in \mathbb{F}} \beta_{1-\varepsilon-\tau_n} \left( \int_{e^n \in \mathcal{E}_1} dP_{Y^n|X^n}(\cdot|x^n) \frac{dP_{E^n}(e^n)}{1 - \tau_n}, Q_Y \right) \quad (81)$$

$$= \inf_{x^n \in \mathbb{F}} \beta_{1-\varepsilon-\tau_n} (P_{Y^n|X^n}(\cdot|x^n), Q_Y). \quad (82)$$

where  $\mathbb{F} = \{x^n : \|x^n\|_2^2 \leq n(\mathbb{E}[E_1] + \delta_n)\}$ . Note that we could take the infimum over  $\mathbb{F}$ , a non-random set here because when  $e^n \in \mathcal{E}_1$ , it implies that  $c(\bar{m}, e^n) \in \mathbb{F}$ . Hence we have (68). ■

Now we are ready to prove the converse bound.

### E. Proof of Theorem 2

From Lemma 2 and 7, we get for any  $\gamma_n > 0$ ,

$$M \leq \frac{\gamma_n}{1 - \varepsilon - \tau_n - \sup_{x^n \in \mathbb{F}} P_{Y^n|X^n=x^n} \left( \frac{dP_{Y^n|X^n=x^n}}{dQ_Y} \geq \gamma_n \right)}. \quad (83)$$

Let  $P_n = \mathbb{E}[E_1] + \delta_n$ . For the EH-AWGN channel, we have that  $P_{Y^n|X^n=x^n} = \mathcal{N}(x^n, \sigma^2 I_n)$  and we pick  $Q_{Y_n} = \mathcal{N}(0, (P_n + \sigma^2)I_n)$ . A few remarks on the choice of  $Q_Y$ .

- We have a lot of flexibility in choosing  $Q_Y$ . However there does exist for every  $n$  an output distribution  $Q_{Y_n}^*$  which gives the tightest upper bound.
- It can be shown that  $\frac{dQ_{Y_n}^*}{dQ_Y} \geq 1$  and, using a change of measure argument, the upper bound for  $Q_{Y_n}^*$  differs from the upper bound for  $Q_Y$  by  $\log \left( \frac{dQ_{Y_n}^*}{dQ_Y} \right)$ .
- For the AWGN channel with equal power constraint  $P$ , the choice of  $Q_Y = \mathcal{N}(0, (P_n + \sigma^2)I_n)$  gives a tight upper bound in the second order terms (See [7]). However we pick this distribution for convenience and that it coincides with the capacity achieving output distribution. Note that if we choose a variance larger or smaller than  $P_n + \sigma^2$ , it will increase the coefficient of the first order term and hence worsen the bound.

From Lemma 5, we obtain the following bound for  $\zeta_n < 0$ .

$$M \leq \frac{2^{nC_n - \zeta_n}}{1 - \varepsilon - \tau_n - \exp \left( \frac{-\zeta_n^2}{2nV_n} \right)} \quad (84)$$

where  $C_n = \frac{1}{2} \log_2(1 + \frac{P_n}{\sigma^2})$  and  $V_n = \frac{P_n(P_n + 2\sigma^2)}{2(P_n + \sigma^2)^2} \log_2^2(e)$ . Let us substitute

$$\zeta_n = -\sqrt{-2nV_n \log(1 - \varepsilon - u_n)} \quad (85)$$

where  $0 < u_n < 1 - \varepsilon$  will be chosen later. This yields

$$\log M \leq nC_n + \sqrt{-2nV_n \log(1 - \varepsilon - u_n)} - \log(u_n - \tau_n). \quad (86)$$

We now observe the following

- $C_n \leq C_{EG} + \frac{\delta_n \log_2(e)}{2(\mathbb{E}[E_1] + \sigma^2)}$ . This follows from using  $(1 + x + y) = (1 + x)(1 + \frac{y}{1+x})$  (for  $x, y \geq 0$ ) followed by  $\ln(1 + x) \leq x$  and
- $V_{EG} \leq V_n \leq V_{EG} + O(\delta_n)$ . This follows from a first order Taylor series expansion of  $V_n$ , where  $\mathbb{E}[E_1]$  is treated as a variable.
- $\log(1 - \varepsilon - u_n) \geq \log(1 - \varepsilon) - \frac{2u_n}{1 - \varepsilon}$  for  $n$  large enough. This is again due to analyzing the first order Taylor expansion.

Gathering all the terms, we get

$$\log M \leq nC_{EG} + \frac{n\delta_n \log_2(e)}{2(E[E_1] + \sigma^2)} + \sqrt{n(V_{EG} + O(\delta_n)) \left( \log \left( \frac{1}{(1-\varepsilon)^2} \right) + \frac{4u_n}{1-\varepsilon} \right)} - \log(u_n - \tau_n). \quad (87)$$

We use the inequality  $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ , to deduce

$$\log M \leq nC_{EG} + \frac{n\delta_n \log_2(e)}{2(E[E_1] + \sigma^2)} + \sqrt{nV_{EG} \log \left( \frac{1}{(1-\varepsilon)^2} \right)} + O(\sqrt{n\delta_n}) + O(\sqrt{nu_n}) - \log(u_n - \tau_n). \quad (88)$$

Now we must pick  $u_n$  and  $\delta_n$ . From Lemma 4, choosing  $u_n = 2\tau_n$ ,  $\delta_n = D_\varepsilon/\sqrt{n}$  where  $D_\varepsilon = \sqrt{\frac{4\sigma_E^2}{1-\varepsilon}}$  (which will imply that  $\tau_n \leq \frac{1-\varepsilon}{4}$ ) gives us

$$\log M \leq nC_{EG} + \frac{\sqrt{n}D_\varepsilon \log_2(e)}{2(E[E_1] + \sigma^2)} + \sqrt{nV_{EG} \log \left( \frac{1}{(1-\varepsilon)^2} \right)} + \sqrt{n(1-\varepsilon)} + O(n^{1/4}). \quad \blacksquare \quad (89)$$

## VI. FINITE BLOCKLENGTH ANALYSIS OF ENERGY HARVESTING DMCs

An energy harvesting DMC is a DMC with an energy harvesting set up at the encoder. Let  $\Lambda(\cdot)$  be the energy function associated with this DMC. The model is the same as that of an EH-AWGN channel except for the following differences and assumptions.

- 1) The AWGN channel is replaced with the DMC.
- 2) Energy consumed by symbol  $x_i$  is  $\Lambda(x_i)$ . Also there is a symbol  $x_0$ , with  $\Lambda(x_0) = 0$ .
- 3) We additionally assume that the DMCs are not exotic<sup>1</sup>(see Appendix H of [7] and also see [8]).

The analysis for energy harvesting DMC's is, by and large, analogous to the analysis of EH-AWGN channels. However, using method of types (refer [14], [16] for more information on types), we are able to improve the converse bound to resemble that of the original non-energy harvesting DMC.

The capacity of an EH-DMC, where the energy harvesting process has mean  $\mathbb{E}[E_1]$ , is given by

$$C_{ED} := \sup_{P \in \mathcal{F}_{\mathbb{E}[E_1]}} I(P; W) \quad (90)$$

where  $\mathcal{F}_a$  was defined in eqn (20).

### A. Energy Harvesting DMC Achievability

We use the same random coding strategy as in the EH-AWGN channel case. Choose any input distribution  $P_X \in \mathcal{F}_{\mathbb{E}[E_1]}$ . Generate an  $M \times n$  matrix with each element distributed i.i.d. with distribution  $P_X$ . Now follow the proof exactly as in the achievability of the EH-AWGN channel case, replacing the term  $X_i^2$  with  $\Lambda(X_i)$  wherever it is encountered. This gives us the following bound on the maximum code size of an EH-DMC which we state as a Theorem.

**Theorem 3.** *Given  $0 < \varepsilon < 1$ , under average probability of error and given the input distribution  $P_X$ , the maximal size of the code  $M^*(\hat{n}, \varepsilon)$  with blocklength  $\hat{n}$  sufficiently large, for an EH-DMC with HUS architecture, and the energy process  $\{E_i\}$  i.i.d. with  $E[E_1^2] < \infty$ , satisfies the following bound*

$$\log M^*(\hat{n}, \varepsilon) \geq \hat{n}I(P_X; W) - \sqrt{\hat{n}}K_\varepsilon I(P_X; W) + \sqrt{\frac{\hat{n}V(P_X; W)}{2}} \Phi^{-1}(\lambda\varepsilon) - \log \hat{n} + O(1). \quad (91)$$

for any  $0 < \lambda < 1$ . Here  $K_\varepsilon = \frac{2\sqrt{\text{Var}(\Delta_1)}}{\mathbb{E}[E_1]\sqrt{(1-\lambda)\varepsilon}}$  and  $\Delta_1 = E_1 - \Lambda(X_1)$ .

In particular, we could substitute  $P_X^* \in \Gamma$  (where  $\Gamma$  is the set of capacity achieving input distributions that are contained in  $\mathcal{F}_{\mathbb{E}[E_1]}$ ) to obtain the best bound. If there are many capacity achieving distributions, then  $V(P_X^*; W)$  may change with the choice of distribution  $P_X^*$ . Hence consider

$$V_{ED} = \begin{cases} V_{\max} := \max_{P \in \Gamma} V(P; W), & \text{if } \varepsilon \leq \frac{1}{2\lambda}, \\ V_{\min} := \min_{P \in \Gamma} V(P; W), & \text{if } \varepsilon > \frac{1}{2\lambda}. \end{cases} \quad (92)$$

Putting it all together, we conclude that the best achievability bound is

$$\log M^*(\hat{n}, \varepsilon) \geq \hat{n}C_{ED} - \sqrt{\hat{n}}K_\varepsilon C_{ED} + \sqrt{\frac{\hat{n}V_{ED}}{2}} \Phi^{-1}(\lambda\varepsilon) - \log \hat{n} + O(1). \quad (93)$$

for all  $\hat{n}$  sufficiently large,

<sup>1</sup>A DMC is exotic if the maximum variance of its information density i.e.  $V_{\max} = 0$  and for some input symbol  $x_0$ ,  $P(x_0) = 0$  for any capacity achieving distribution  $P$  but  $D(W(\cdot|x_0)||Q_Y^*) = C$  and  $V(W(\cdot|x_0)||Q_Y^*) > 0$ .

### B. Finite Blocklength Converse for EH-DMC

To obtain a good converse bound for EH-DMC, we will be using concepts from method of types. Refer Appendix A for notations. Let the DMC of the EH-DMC be given by the random transformation  $W(y|x)$ . Denote

$$v_n := \mathbb{E}[E_1] + \delta_n. \quad (94)$$

Recall the modified meta converse in lemma 7.

$$\log M^*(n, \varepsilon) \leq -\log \left\{ \sup_{\mathbf{x} \in \mathbb{F}} \beta_{1-\varepsilon-\tau_n}(W^n(\cdot|\mathbf{x}), Q_{\mathbf{Y}}) \right\}. \quad (95)$$

Consider the following term:

$$\sup_{\mathbf{x} \in \mathbb{F}} \beta_{1-\varepsilon-\tau_n}(W^n(\cdot|\mathbf{x}), Q_{\mathbf{Y}}). \quad (96)$$

Note that  $\mathbb{F} = \mathbb{F}_{v_n} = \{\mathbf{x} : \|\mathbf{x}\|^2 \leq nv_n\}$  here. According to the type transformation trick (TTT) (Lemma 8 in Appendix A), this is equivalent to

$$\sup_{P \in \mathcal{F}_{v_n} \cap \mathcal{P}_n} \sup_{\mathbf{x} \in \mathcal{T}_P} \beta_{1-\varepsilon-\tau_n}(W^n(\cdot|\mathbf{x}), Q_{\mathbf{Y}}). \quad (97)$$

Now consider the inner supremum term,

$$\sup_{\mathbf{x} \in \mathcal{T}_P} \beta_{1-\varepsilon-\tau_n}(W^n(\cdot|\mathbf{x}), Q_{\mathbf{Y}}). \quad (98)$$

It actually holds that the beta error function above is independent of which sequence  $\mathbf{x}$  is picked provided that the sequences have the same type and  $Q_{\mathbf{Y}} = \prod_{k=1}^n Q_Y$  for some distribution  $Q_Y$  on  $\mathcal{Y}$ . Hence pick any sequence  $\mathbf{x}$  from  $\mathcal{T}_{P_0}$  where  $P_0 \in \mathcal{F}_{v_n}$ . Note that this argument was used in proving Theorem 48 in [7] while applying the meta converse to a standard DMC. We show that the arguments in [7] are applicable here with only a few minor changes.

Let  $Q_Y = P_0 W$ . We recall Theorem 48 from [7] for standard, non-exotic DMCs. Although the bound was originally for the maximal subcode of type  $P_0$  of the maximal code, we note that the term actually being bounded is the beta error function as mentioned below.

**Theorem 4.** For  $0 < \varepsilon < 1$ , for all  $P_0 \in \mathcal{P}_n$ ,  $\mathbf{x} \in \mathcal{T}_{P_0}$  and  $n$  sufficiently large, we have

$$-\log \beta_{1-\varepsilon}(W^n(\cdot|\mathbf{x}), (P_0 W)^n) \leq nC_D + \sqrt{nV_D} \Phi^{-1}(\varepsilon) + \frac{1}{2} \log n + O(1) \quad (99)$$

where

$$V_D = \begin{cases} V_{\min} = \min_{P \in \Gamma} V(P; W), & 0 < \varepsilon \leq 1/2, \\ V_{\max} = \max_{P \in \Gamma} V(P; W), & 1/2 < \varepsilon < 1, \end{cases} \quad (100)$$

where  $\Gamma$  is the set of capacity achieving distributions.

Note that the bound on RHS does not depend on the distribution of the type. Hence if we make the following substitutions

- 1) Replace  $\Gamma$  with

$$\Gamma_{v_n} = \{P \in \mathcal{F}_{v_n} : I(P; W) = C\} \quad (101)$$

This is because the outer supremum in (97) is over  $\mathcal{F}_{v_n}$ . Note that the original proof of Theorem 4 used the fact that  $\Gamma$  was compact and convex. These properties hold for  $\Gamma_{v_n}$  so we may substitute this wherever  $\Gamma$  was used.

- 2) The final supremum that gives the uniform (over input distributions) bound was over  $\mathcal{P}$ . Here we substitute  $\mathcal{F}_{v_n}$  in its place.
- 3)  $\varepsilon$  is replaced by  $\varepsilon + \tau_n$ .

Putting all these facts together gives us

$$\log M^*(n, \varepsilon) \leq nC_D(v_n) + \sqrt{n\hat{V}(v_n)} \Phi^{-1}(\varepsilon + \tau_n) + O(\log(n)), \quad (102)$$

where

$$\hat{V}(v_n) = \begin{cases} V_{\min}^{(n)} = \min_{P \in \Gamma_{v_n}} V(P; W), & 0 < \varepsilon + \tau_n \leq 1/2, \\ V_{\max}^{(n)} = \max_{P \in \Gamma_{v_n}} V(P; W), & 1/2 < \varepsilon + \tau_n < 1. \end{cases} \quad (103)$$

We can further simplify (102) by expanding  $C_D(v_n)$ ,  $\hat{V}(v_n)$  and  $\Phi^{-1}(u)$ .

Now  $C_D(a)$  is a non-decreasing concave function (see [16]). Hence we have for any  $a > 0, b > 0$ ,

$$C_D(a+b) \leq C_D(a) + bC'_D(a), \quad (104)$$

where  $C'_D(\cdot)$  is the derivative of  $C_D(a)$ . Let  $a = \mathbb{E}[E_1]$  and  $b = \delta_n$ . Note that  $C'_D(a)$  in this case is a constant since  $\mathbb{E}[E_1]$  is a constant.

Using Taylor series approximation, we get that for some constant  $K_\varepsilon$ ,

$$\Phi^{-1}(\varepsilon + \tau_n) \leq \Phi^{-1}(\varepsilon) + \tau_n K_\varepsilon. \quad (105)$$

We will ensure  $\tau_n \leq \frac{1-\varepsilon}{4}$ . Hence, pick  $\delta_n = \frac{D_\varepsilon}{\sqrt{n}}$  where  $D_\varepsilon = \sqrt{\frac{4\sigma_E^2}{1-\varepsilon}}$ . Now let  $\varepsilon_R$  be the root of

$$\Phi^{-1}(\varepsilon) + \frac{K_\varepsilon(1-\varepsilon)}{4} = 0. \quad (106)$$

Pick any  $\eta > 0$ . Observe that for  $n$  sufficiently large,  $\Gamma_{v_n} \subset \Gamma_{\mathbb{E}[E_1] + \eta}$ . Hence we can replace  $\hat{V}(v_n)$  with

$$V_\varepsilon^*(\eta) = \begin{cases} \min_{P \in \Gamma_{\mathbb{E}[E_1] + \eta}} V(P; W), & 0 < \varepsilon \leq \varepsilon_R, \\ \max_{P \in \Gamma_{\mathbb{E}[E_1] + \eta}} V(P; W), & \varepsilon_R < \varepsilon < 1. \end{cases} \quad (107)$$

Note that  $C_D(\mathbb{E}[E_1]) \equiv C_{ED}$ , Thus we have for  $n$  sufficiently large

$$\log M^*(n, \varepsilon) \leq nC_{ED} + \sqrt{n}C'(\mathbb{E}[E_1])D_\varepsilon + \sqrt{nV_\varepsilon^*(\eta)} \left( \Phi^{-1}(\varepsilon) + \frac{K_\varepsilon(1-\varepsilon)}{4} \right) + O(\log(n)). \quad (108)$$

Observe that unlike the AWGN case, the second coefficient of  $\sqrt{n}$  can actually be negative and this gives a better converse than that of EH-AWGN channels.

## VII. CONCLUSION AND FUTURE WORK

We have shown that for an EH-AWGN channel and EH-DMC with infinite energy buffer and i.i.d. energy arrivals, the maximum code size  $M$  varies as  $\log M = nC_{EG} + O(\sqrt{n})$ . Comparing the achievability bound with that of a standard AWGN channel with power constraint  $\mathbb{E}[E_1]$ , we conjecture that energy harvesting hurts the second order term. It is a conjecture as the converse and achievability do not have matching bounds in second order terms. Obtaining matching bounds and quantifying the effect of an energy harvesting system on higher order terms could be topics for future research.

## ACKNOWLEDGEMENTS

The authors would like to thank Himanshu Tyagi and Deekshith P.K. for their valuable inputs and helpful discussions while preparing this manuscript.

## APPENDIX A METHOD OF TYPES

The method of types or empirical distributions is a powerful tool to analyze discrete memoryless channels and in source coding. We only recall the definitions and results that we will use in Section VI-B. Refer [16], [14] for more details.

Consider a finite discrete alphabet  $\mathcal{X}$ . Given an  $n$  length sequence  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ . We define the *type* of  $\mathbf{x}$  as the probability distribution

$$P_{\mathbf{x}}(x) = \frac{\sum_{k=1}^n 1_{\{x_k=x\}}}{n} = \frac{N(x|\mathbf{x})}{n} \quad \forall x \in \mathcal{X}, \quad (109)$$

where  $N(x|\mathbf{x})$  is the number of times the symbol  $x$  appears in  $\mathbf{x}$ . Note that multiple sequences can have the same type. More importantly, these sequences are permutations of one another. Define the type class  $\mathcal{T}_P$  as the set of  $n$  length sequences with type  $P$ . Finally we denote the set of all types of  $n$  length sequences by  $\mathcal{P}_n$ . Note that not every distribution is a type, but  $\cup_{n \geq 1} \mathcal{P}_n$  is dense in  $\mathcal{P}(\mathcal{X})$ .

We now introduce the type transformation trick (TTT) which when applied to a function of sequences gives rise to a function of types.

**Lemma 8 (TTT).** *Let  $g : \mathcal{X} \rightarrow \mathbb{R}^+$ . Then suppose a sequence  $\mathbf{x}$  is of type  $P$ , we have*

$$\sum_{k=1}^n g(x_k) = n\mathbb{E}_P[g(X)]. \quad (110)$$

Now consider the sets  $\mathbb{F}_a$  and  $\mathcal{F}_a$  as defined in (19) and (20) respectively. Let  $h : \mathcal{X}^n \rightarrow \mathbb{R}$  be a function. Then the following holds

$$\sup_{\mathbf{x} \in \mathbb{F}_a} h(\mathbf{x}) = \sup_{P \in \mathcal{F}_a} \sup_{\mathbf{x} \in \mathcal{T}_P} h(\mathbf{x}) \quad (111)$$



The proof of the first statement is a simple exercise. For the second part, observe that

$$\mathbf{x} \in \mathbb{F}_a \cap \mathcal{T}_P \iff \sum_{k=1}^n \Lambda(x_k) \leq na, \mathbf{x} \in \mathcal{T}_P \quad (112)$$

$$\iff \mathbb{E}_P[\Lambda(X)] \leq a, \mathbf{x} \in \mathcal{T}_P \quad (113)$$

The first statement in (113) is a restriction on allowed probability distributions which when we intersect with  $\mathcal{P}_n$  gives us  $\mathcal{F}_a$ .

## APPENDIX B

### INVARIANCE OF BETA ERROR FUNCTION TO SEQUENCES WITH SAME TYPE

Let us fix a type  $P \in \mathcal{P}_n$  for some  $n \geq 1$ . Consider two  $n$  length sequences  $\mathbf{x}_1$  and  $\mathbf{x}_2$  with type  $P$ . Then there exists a permutation  $\pi$  such that  $\pi(\mathbf{x}_2) = \mathbf{x}_1$ . Denote  $\pi_i(\mathbf{x})$  as the  $i$ th coordinate of  $\pi(\mathbf{x})$ . Consider a DMC  $W$  and a product distribution  $Q_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^n Q_Y(y_i)$ . Then there exists an admissible test function  $T_1$  such that

$$\beta_\alpha(W^n(\cdot|\mathbf{x}_1), Q_{\mathbf{Y}}) = \sum_{\mathbf{y}} Q_{\mathbf{Y}}(\mathbf{y}) P_{T_1|\mathbf{Y}}(1|\mathbf{y}) = \sum_{\mathbf{y}} Q_{\mathbf{Y}}(\mathbf{y}) P_{T_1|\mathbf{Y}}(1|\pi(\mathbf{y})), \quad (114)$$

where the last equality follows by observing that the sum does not change if we change the order of summation. The product does not change if we permute the indices either. That is,

$$Q_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^n Q_Y(y_i) = \prod_{i=1}^n Q_Y(\pi_i(\mathbf{y})) = Q_{\mathbf{Y}}(\pi(\mathbf{y})). \quad (115)$$

Now we have by definition of an admissible test function

$$\alpha \leq \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}_1) P_{T_1|\mathbf{Y}}(1|\mathbf{y}) \quad (116)$$

$$= \sum_{\mathbf{y}} W^n(\pi(\mathbf{y})|\mathbf{x}_1) P_{T_1|\mathbf{Y}}(1|\pi(\mathbf{y})) \quad (117)$$

$$= \sum_{\mathbf{y}} \prod_{i=1}^n W(\pi_i(\mathbf{y})|x_{1i}) P_{T_1|\mathbf{Y}}(1|\pi(\mathbf{y})) \quad (118)$$

$$= \sum_{\mathbf{y}} \prod_{i=1}^n W(y_i|\pi_i^{-1}(\mathbf{x}_1)) P_{T_1|\mathbf{Y}}(1|\pi(\mathbf{y})) \quad (119)$$

$$= \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}_2) P_{T_1|\mathbf{Y}}(1|\pi(\mathbf{y})). \quad (120)$$

Therefore for every admissible test  $T_1$  for sequence  $\mathbf{x}_1$ , there exists a test  $T_2$  for sequence  $\mathbf{x}_2$  which is admissible and is given by

$$P_{T_2|Y^n}(1|\mathbf{y}) = P_{T_1|\mathbf{Y}}(1|\pi(\mathbf{y})) \quad \forall \mathbf{y} \in \mathcal{Y}^n. \quad (121)$$

Hence we get

$$\beta_\alpha(W^n(\cdot|\mathbf{x}_2), Q_{\mathbf{Y}}) \leq \sum_{\mathbf{y}} Q_{\mathbf{Y}}(\mathbf{y}) P_{T_2|Y^n}(1|\mathbf{y}) \quad (122)$$

$$= \sum_{\mathbf{y}} Q_{\mathbf{Y}}(\mathbf{y}) P_{T_1|\mathbf{Y}}(1|\pi(\mathbf{y})) \quad (123)$$

$$= \beta_\alpha(W^n(\cdot|\mathbf{x}_1), Q_{\mathbf{Y}}). \quad (124)$$

By interchanging  $\mathbf{x}_1$  and  $\mathbf{x}_2$  and the corresponding terms in the above proof, we can show the reverse inequality. Hence

$$\beta_\alpha(W^n(\cdot|\mathbf{x}_2), Q_{\mathbf{Y}}) = \beta_\alpha(W^n(\cdot|\mathbf{x}_1), Q_{\mathbf{Y}}). \quad (125)$$

Therefore for all  $\mathbf{x} \in \mathcal{T}_P$ ,  $\beta_\alpha(W^n(\cdot|\mathbf{x}), Q_{\mathbf{Y}})$  is invariant to the choice of  $\mathbf{x}$ .

## REFERENCES

- [1] K. G. Shenoy and V. Sharma, "Finite blocklength achievable rates for energy harvesting awgn channels with infinite buffer," *IEEE International Symposium on Information Theory*, 2016.
- [2] V. Potdar, A. Sharif, and E. Chang, "Wireless sensor networks: A survey," in *Advanced Information Networking and Applications Workshops, 2009. WAINA'09. International Conference on*. IEEE, 2009, pp. 636–641.
- [3] H. Erkal, F. M. Ozcelik, M. A. Antepli, B. T. Bacinoglu, and E. Uysal-Biyikoglu, "A survey of recent work on energy harvesting networks," in *Computer and Information Sciences II*. Springer, 2011, pp. 143–147.
- [4] A. Kumar, K. Singh, and D. Bhattacharya, "Green communication and wireless networking," in *Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on*. IEEE, 2013, pp. 49–52.
- [5] V. Strassen, "Asymptotische abschätzungen in shannons informationstheorie," in *Trans. Third Prague Conf. Inf. Theory*, 1962, pp. 689–723.
- [6] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.
- [7] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [8] M. Tomamichel and V. Y. Tan, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7041–7051, 2013.
- [9] —, "Second-order coding rates for channels with state," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4427–4448, 2014.
- [10] R. Rajesh, V. Sharma, and P. Viswanath, "Capacity of gaussian channels with energy harvesting and processing cost," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2563–2575, 2014.
- [11] O. Ozel and S. Ulukus, "Achieving awgn capacity under stochastic energy harvesting," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6471–6483, 2012.
- [12] J. Yang, "Achievable rate for energy harvesting channel with finite blocklength," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 811–815.
- [13] S. L. Fong, V. Y. Tan, and J. Yang, "Non-asymptotic achievable rates for energy-harvesting channels using save-and-transmit," *IEEE Journal on Selected Areas in Communications*, no. 99, 2015.
- [14] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
- [15] T. Han, "Information-spectrum methods in information theory [english translation]. series: Stochastic modelling and applied probability, vol. 50," Springer, vol. 1, no. 6, pp. 3–1, 2003.
- [16] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
- [17] V. Kostina and S. Verdú, "Channels with cost constraints: strong converse and dispersion," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2415–2429, 2015.
- [18] K. B. Athreya and S. N. Lahiri, *Probability Theory*. Texts and Readings in Mathematics, Hindustan Book Agency, 2006.
- [19] C. E. Shannon, "Certain results in coding theory for noisy channels," *Information and control*, vol. 1, no. 1, pp. 6–25, 1957.
- [20] I. S. Tyurin, "An improvement of upper estimates of the constants in the lyapunov theorem," *Russian Mathematical Surveys*, vol. 65, no. 3, pp. 201–202, 2010.
- [21] L. Birgé, "An alternative point of view on lepsi's method," *Lecture Notes-Monograph Series*, pp. 113–133, 2001.